

Figure S1: Data analysis pipeline for Differential Gene Expression analysis applied on the ITAN RNASeq dataset.

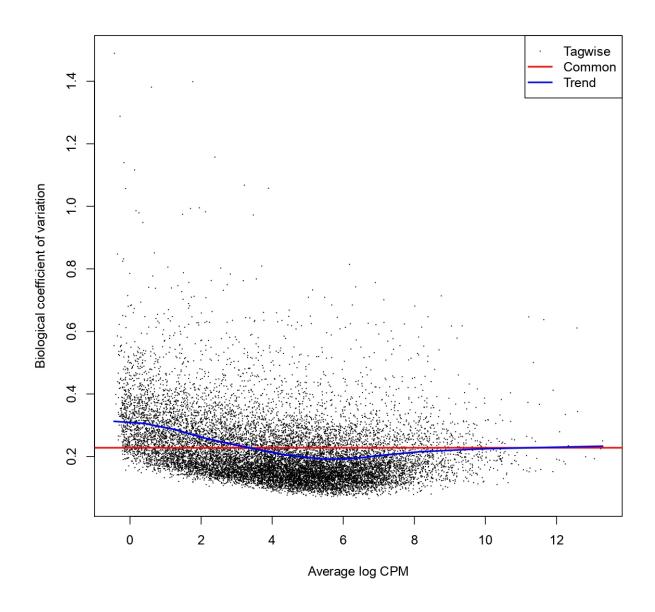


Figure S2: Estimation of the biological coefficient of variation used in the generalized linear model for DGE implemented in edgeR package.

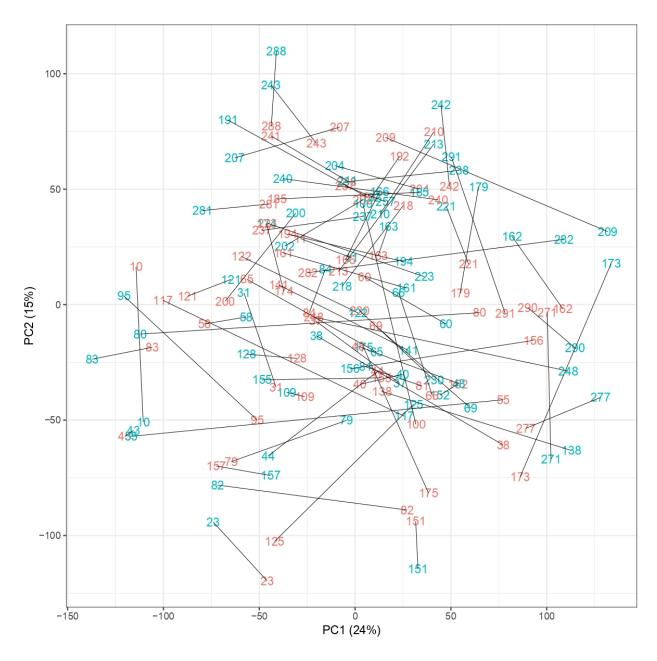


Figure S3: Principal Component Analysis on normalized gene counts, (Blue=ASD sibling, Red=Controls sibling, lines connecting siblings).

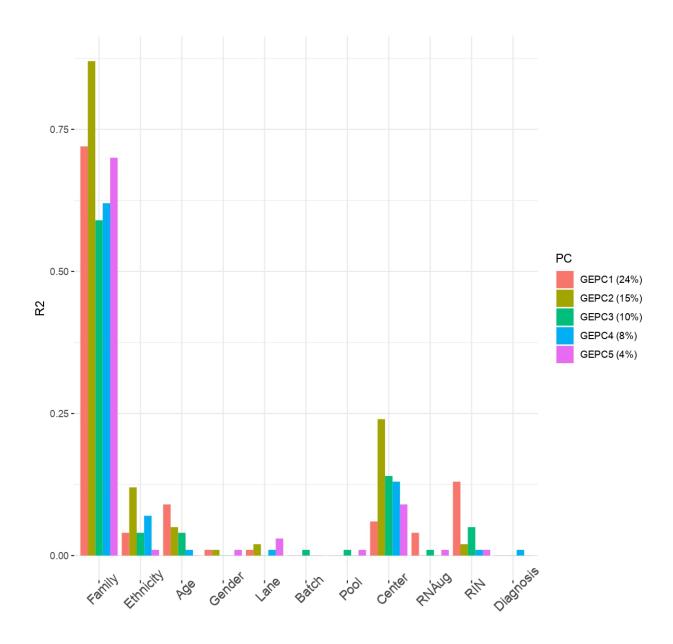


Figure S4: Variance explained by each available covariate for each of the first 5 principal components.

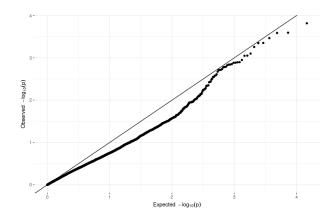


Figure S5a: DGE analysis on male samples only (with deconvolution covariate)

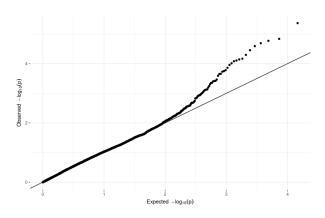


Figure S5b: DGE analysis on male samples only (no cell deconvolution covariate)

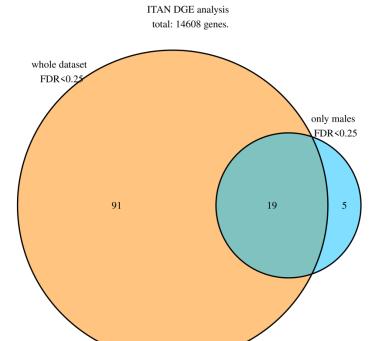


Figure S6: Venn diagram of the intersection between male DGE and whole dataset

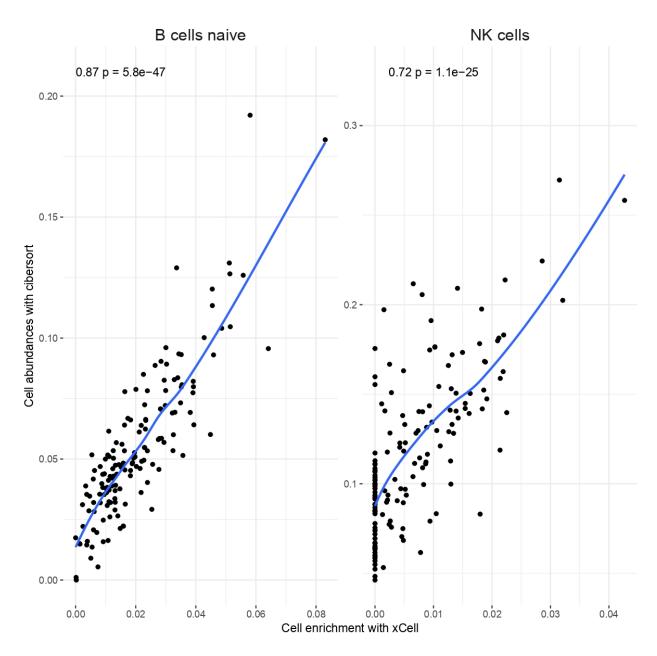
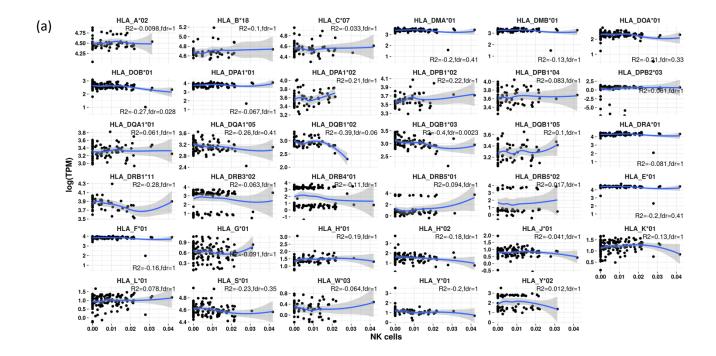


Figure S7: Correlation between cell estimates derived from different methods for cell deconvolution.



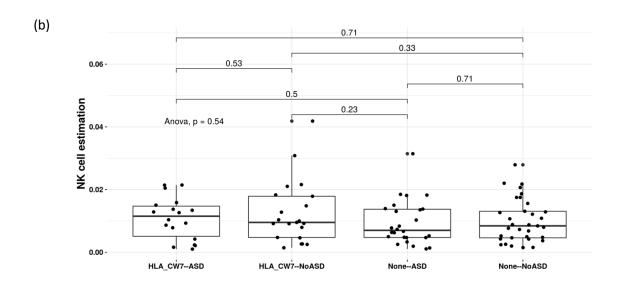


Figure S8: (a) Expression of HLA-allele transcripts (Transcripts Per Million, TPM) and correlation with xCell estimated NK cell enrichment. Only transcripts present in at least 30% of the samples are reported in the plots. (b) NK cell estimates within diagnostic categories based on presence/absence of HLA-Cw7. None of the comparisons have a significant p.value (p<0.05). Samples with a NK cell estimation score lower than 1e-6 have been removed from the plot.

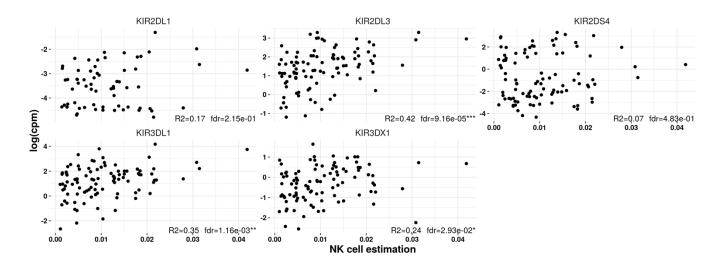


Figure S9: Correlation of expression (in log Count Per Million, CPM) for KIR genes from RNASeq and xCell enrichment score for NK cell abundance. Only transcripts with CPM > 1 in at least 30% of the samples were included in the plot. The significant correlation found for KIR2DL3, KIR3DL1 and KIR3DX1 supports the notion of a positive regulatory effect of KIR genes on NK cells

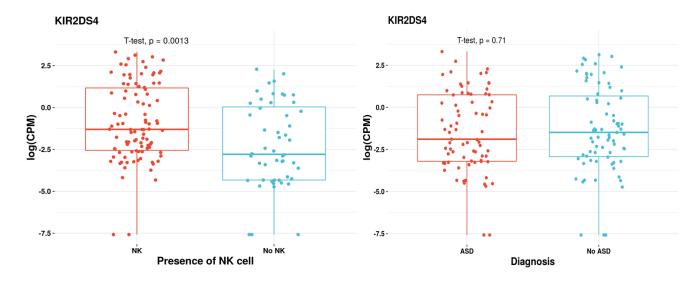


Figure S10: KIR2DS4 expression in log Count Per Million reads (CPM) estimated by RNASeq grouped by presence/absence of NK cell (based on a threshold of NK cell enrichment score > 10E-06), left panel. The difference observed support the NK cell activating role of KIR2DS4. Nevertheless, KIR2DS4 expression is not significantly different between ASD cases and unaffected controls (right panel).

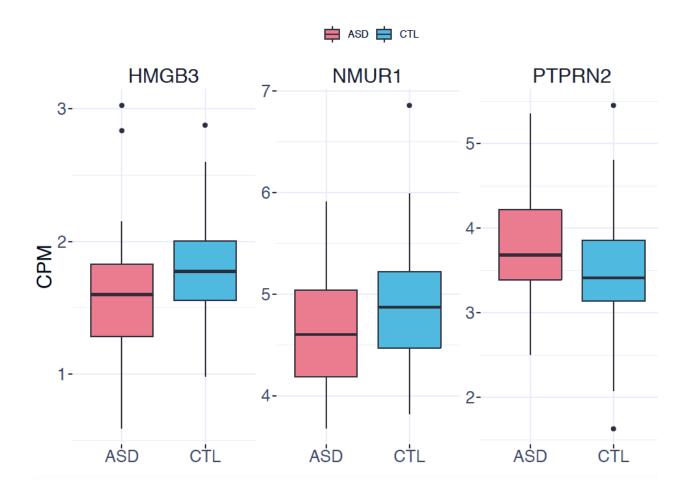


Figure S11: Gene expression signature differences between discordant siblings for statistically significant genes (at FDR<0.25) after accounting for cell composition.

Pvalue-method			HMGB3	Es-method	Es-method				
TestStatistic	p_value	FDR		TestStatistic	p_value	FDR			
3.55	0.00	0.03		3.45	0.00	0.05			

Study	Total	Mean	Case SD		Co Mean	ntrol SD		Standa dif	rdised ferenc		SMD	95%-CI	Weight
Chow 2012 male cortex Voineagu 2011 male cortex Ginsberg 2012 male cortex		0.00	0.36	25		0.24		-	‡		-1.04	[-1.21; 0.19] [-1.71; -0.37] [-2.30; 0.44]	42.7% 46.2% 11.1%
Random effects model Heterogeneity: $I^2 = 0\%$ , $\tau^2 = 0$ , $\rho$		5		49			-2	-1	0	1	 -0.80	[-1.26; -0.35]	100.0%

Study	Total Me	Case an SD Total	Control Mean SD	Standardised mean difference	SMD	95%-CI Weight
Voineagu 2011 cerebellum Ginsberg 2012 cerebellum			7.11 0.27 7.17 0.23			[-1.30; 0.83] 48.4% [-1.81; 0.25] 51.6%
Random effects model Heterogeneity: $I^2 = 0\%$ , $\tau^2 = 0$ ,	<b>13</b> p = 0.47	19		-1.5 -1 -0.5 0 0.5 1 1.5	-0.52	[-1.25; 0.22] 100.0%

Study	Total	C Mean	ase SD	Total	Co Mean	ntrol SD		Standardised mean difference	SMD	95%-CI	Weight
Chow 2012 male cortex Voineagu 2011 male cortex Ginsberg 2012 male cortex Chow 2012 female cortex Voineagu 2011 female cortex Chow 2012 fixed cortex	15 16 4 4 7 4	8.88 ( 9.70 1 9.79 (	0.36 0.38 1.02 0.66		10.11 10.07	0.24 0.28 0.19 0.34		*	-0.51 -1.04 -0.93 -0.49 -0.40 0.30	[-1.21; 0.19] [-1.71; -0.37] [-2.30; 0.44] [-1.91; 0.94] [-1.99; 1.19] [-0.83; 1.44]	31.8% 34.4% 8.2% 7.6% 6.1% 11.9%
Random effects model Heterogeneity: $I^2 = 0\%$ , $\tau^2 = 0$ , $p$	<b>50</b> = 0.49			67			-2	-1 0 1	 -0.62	[-1.01; -0.23]	100.0%

Figure S12: Expression profiles of HMGB3 in brain tissue from human ASD studies extracted from dbMEGA (meta-analysis of ASD databases, <a href="https://dbmdega.shinyapps.io/dbMDEGA/">https://dbmdega.shinyapps.io/dbMDEGA/</a>, Zhang et al., 2017)

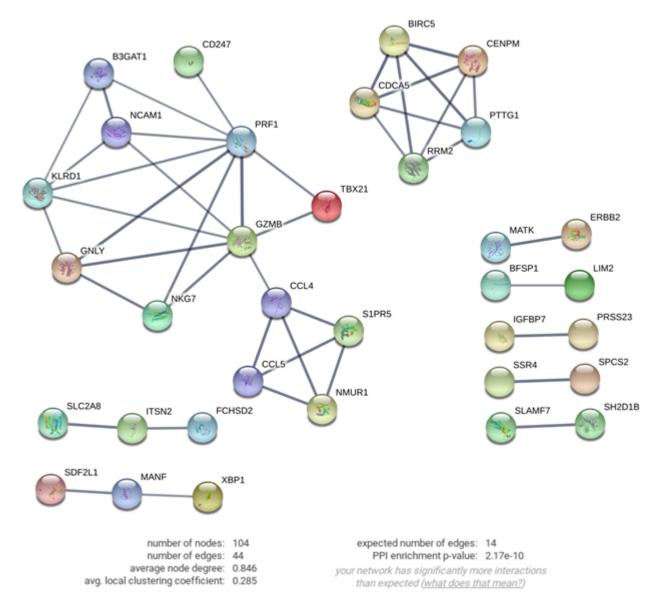


Figure S13: Enrichment on STRING protein-protein interaction database of the list of differentially expressed genes with FDR <0.25. All available interaction sources have been used to infer confidence of interaction between the submitted genes. Only links with a confidence of 0.7 or more are shown in the graph. Nodes without connection have been removed from the plot.